

Géométriser le sens lexical

La synonymie comme accès à la sémantique

Fabienne Venant

Loria / Université Nancy2
615, rue du Jardin Botanique
54600 Villers lès Nancy Cedx
fabienne.venant@loria.fr

Abstract

Cet article présente une modélisation du sens lexical, centrée sur le phénomène de la polysémie. Ce modèle utilise la relation de synonymie comme accès aux informations lexico-sémantiques, et propose une réponse géométrique à la question de la représentation du sens. L'étude approfondie des relations de synonymie permet de mettre en évidence à la fois le fonctionnement des unités polysémiques prises individuellement, et leur place dans l'organisation globale du lexique. Il s'agit donc de construire des espaces sémantiques à différentes échelles, pouvant rendre compte de la sémantique d'une unité donnée ou permettre un accès à la structuration sémantique du lexique. La ressource ainsi construite permet d'accéder de façon synthétique à la sémantique des entrées d'un dictionnaire de synonymes. Des pistes sont envisagées vers une exploration plus globale du lexique ainsi que vers une articulation des niveaux conceptuels et sémantiques.

1 Introduction

Le travail présenté ici s'inscrit dans un cadre plus général de modélisation des processus de construction du sens au sein d'un énoncé (Victorri & Fuchs, 1996, Venant, 2006). Ces travaux nous ont amenés à accorder une place centrale aux phénomènes de la polysémie et de la synonymie. Ces phénomènes touchent une grande diversité d'unités linguistiques. Tenter de les formaliser suppose d'une part de définir le sens lexical, et la représentation que l'on veut en donner, et d'autre part de rendre compte des relations sémantiques entre les unités lexicales. Nous avons tenté de donner une réponse géométrique à ces questions, à travers la relation de synonymie. Nous associons donc à chaque vocable polysémique un espace sémantique continu, dans lequel les sens des différentes acceptions s'organisent selon leurs proximités sémantiques. Ces espaces sont construits automatiquement, pour n'importe quelle entrée du Dictionnaire Electronique des Synonymes (DES²). Le DES est pour cela modélisé sous forme de graphe. L'algorithme de construction des espaces sémantiques repose sur l'analyse de ce graphe.

2 Sens lexical

Le modèle présenté ici est un modèle continu du sens lexical. Notons que quand nous parlons de continu, il s'agit bien de caractériser le modèle du sens utilisé et non le sens lui-même. Comme le remarque Polguère (2002): « C'est un lieu commun de dire qu'il n'existe pas de consensus sur ce qu'il faut entendre par sens linguistique. Toute définition de cette notion ne peut qu'être partielle en regard des différentes façons d'aborder l'étude du contenu des énoncés. » Nous travaillons ici dans le cadre de la construction dynamique du sens (Victorri & Fuchs, 1996) qui considère que le sens d'une unité linguistique dans un énoncé

² <http://www.crisco.unicaen.fr/cgi-bin/cherches.cgi>.

donné est le résultat de l'interaction entre un apport sémantique constant associé à cette unité, que l'on peut appeler son *noyau de sens*, ou encore sa *forme schématique*, et le contexte d'énonciation de cette unité. Le contexte d'énonciation comporte à la fois les autres unités linguistiques présentes dans l'énoncé (son cotexte) et la situation extralinguistique dans laquelle cet énoncé est proféré. Il ne s'agit pas de nier le fait que l'expression linguistique fournisse une composante du matériel de base requis pour l'élaboration du sens, mais de souligner le fait qu'elle n'est qu'une parmi ces composantes. Cette position peut-être rapprochée de celle de Croft et Cruse (2004). Pour ces auteurs, les sens sont des choses qu'on élabore en utilisant les propriétés des éléments linguistiques comme indices partiels tout au long de connaissances non linguistiques, mais aussi les informations disponibles à partir du contexte et les connaissances ou conjectures quant à l'état d'esprit du locuteur.

Le noyau de sens n'est donc pas un sens à proprement parler, mais plutôt un schéma de base à partir duquel se construisent les différents sens d'une unité. Pour l'appréhender, il faut tenter de cerner les propriétés du mot lui-même, qui expliquent à la fois qu'il puisse prendre des sens différents selon les énoncés, et en même temps qu'il évoque à lui tout seul la classe d'objets à laquelle on l'associe spontanément. Pour en donner une idée plus intuitive, disons que le noyau de sens est précisément ce qu'on cherche à appréhender lorsqu'en consultant un dictionnaire, on parcourt l'ensemble de l'article concernant un mot pour se faire une idée de la façon dont il « fonctionne » en contexte.

Prenons par exemple le nom *cours*. L'examen des différentes définitions constituant l'article associé à ce nom dans le Trésor de la Langue Française Informatisé (TLFI)³, nous permet de dégager le noyau de sens suivant :

Développement (écoulement) continu entre des limites précises, qui peut ensuite se décliner selon trois perspectives :

- une perspective spatiale : *cours d'une rivière, le cours du sang dans notre corps, cours du soleil, cours de Vincennes* ;
- une perspective temporelle : (temps lui-même) *cours des saisons*, (suite de choses ou d'événements), *cours des pensées*, (valeurs) *cours du coton* ;
- une perspective que le TLFI qualifie de spatio-temporelle, mais qu'on pourrait plutôt qualifier de notionnelle : *cours de latin, donner, faire, suivre un cours*, par métonymie, traité, manuel, établissement - (en parlant d'autres activités) *au long cours*.

En contexte, ce noyau de sens entre en interaction avec les autres éléments linguistiques pour donner naissance à un sens précis. Les énoncés 1 et 2 illustrent ce phénomène. Dans l'énoncé 1, l'interaction entre les noyaux de sens de *cours*, de *rivière* et de *suivre* conduit à la sélection de la perspective spatiale. L'énoncé 2, quant à lui, met en jeu la perspective spatio-temporelle de *cours*.

(1) *La rivière suit son cours.*

(2) *L'étudiant suit un cours.*

3 Des espaces sémantiques pour représenter le sens lexical

Il s'agit ici d'associer à chaque vocable polysémique un espace continu qui rende compte de son noyau de sens. On veut capturer et représenter son organisation sémantique : les différentes nuances de sens qu'il peut prendre, leurs proximités sémantiques, comment on peut passer continûment de l'une à l'autre. L'idée est d'utiliser la relation de synonymie comme accès à cette structure sémantique.

Les espaces sémantiques sont construits automatiquement à partir d'un graphe de synonymie tiré du DES. La méthode utilisée est celle initialement proposée par Ploux et Victorri (1998). On explore ce dictionnaire sous la forme d'un graphe : les sommets sont les expressions linguistiques. Un lien est tracé entre deux expressions lorsque le dictionnaire signale qu'elles sont synonymes. L'idée sous-jacente à la construction des espaces sémantiques est qu'un synonyme n'est généralement pas suffisant pour définir un sens précis d'une expression. Ainsi, au sein de sa synonymie avec *cours*, *fil* est à la fois synonyme de

³ <http://atilf.atilf.fr/tlf.htm>.

courant et de *déroulement*, ce qui correspond à deux sens distincts de *cours*, l'un lié à la perspective spatiale (ou notionnelle), l'autre à la perspective temporelle. Or les points de l'espace sémantique doivent correspondre à des sens précis de l'unité. C'est pourquoi nous avons recours à la notion de clique. Une clique est un sous-graphe complet maximal, c'est-à-dire un ensemble de sommets, le plus grand possible, reliés deux à deux. Voici par exemple la liste des cliques qui, dans le sous-graphe formé par *cours* et tous ses synonymes, contiennent le nom *fil*. Chaque clique correspond à une nuance possible de sens pour *cours* : < *courant* ; *cours* ; *fil* ; *flot* >, < *courant* ; *cours* ; *fil* ; *succession* > et < *cours* ; *déroulement* ; *enchaînement* ; *fil* ; *succession* ; *suite* >

L'espace sémantique est une projection en deux dimensions du nuage formé par les cliques dans l'espace multidimensionnel engendré par les synonymes de l'expression considérée. Il est muni de la métrique du chi2. C'est elle qui s'est en effet avérée la plus efficace pour obtenir une représentation respectant la notion intuitive de proximité entre sens. La Figure 1 présente l'espace sémantique de *cours*.

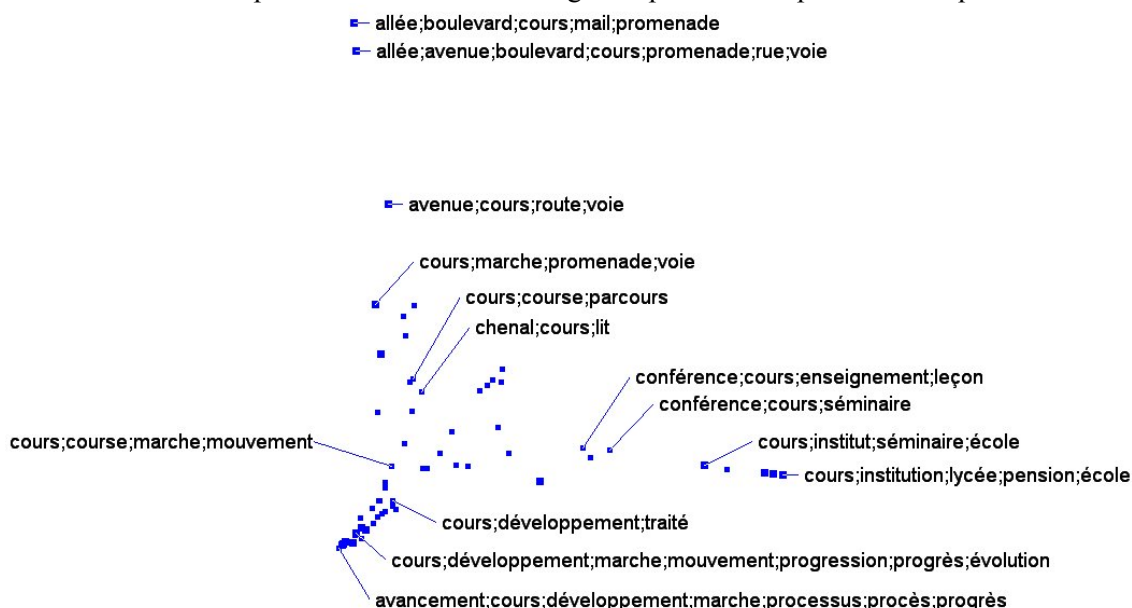


Figure 1. Espace sémantique associé à *cours*

4 Continuum de sens

Les espaces sémantiques ainsi construits permettent de rendre compte d'un continuum entre les différents sens d'une unité, et par là même d'appréhender le noyau de sens de cette unité. On voit par exemple, sur la Figure 1, que l'espace sémantique de *cours* fait clairement apparaître deux branches, l'une (partie haute de l'espace sémantique) correspondant à la perspective spatiale, avec des cliques comme < *allée* ; *boulevard* ; *cours* ; *mail* > ou < *chenal* ; *cours* ; *voie* >, l'autre (vers la droite) correspondant à la perspective notionnelle, avec des cliques comme < *conférence* ; *cours* ; *séminaire* > ou < *cours* ; *mouvement* ; *école* >. On passe continûment d'une branche à l'autre, via la zone centrale, qui rassemble des cliques relevant plutôt de la perspective temporelle comme < *cours* ; *durée* ; *courant* > ; ou < *cours* ; *fil* ; *succession* >. Une telle représentation permet, entre autres, d'expliciter la relation de famille unissant deux sens d'une unité donnée. Ainsi on peut ainsi exhiber une suite de sens intermédiaires permettant de passer continûment du sens de *cours* dans « la rivière suit son cours » à celui qu'il prend dans un étudiant « L'étudiant suit un cours », via les cliques suivantes :

- < *chenal* ; *lit* ; *cours* >
- < *chenal* ; *courant* ; *cours* >
- < *cours* ; *course* ; *marche* ; *mouvement* >
- < *courant* ; *cours* ; *flot* ; *mouvement* >

< courant ; cours ; fil ; succession >
 < courant ; cours ; mouvement >
 < cours ; mouvement ; école >
 < cours ; classe ; leçon ; école >
 < cours ; enseignement ; leçon ; conférence >

5 Caractériser la polysémie

Partant du fait que le sens d'une expression linguistique dans une occurrence donnée dépend en partie de ce qu'apporte cette expression elle-même de constant, quel que soit le contexte, et en partie de ce qui est fonction du contexte, on peut classer les expressions suivant l'importance relative de ces deux facteurs. A un extrême, le contexte ne joue aucun rôle : l'expression est monosémique ; son sens est le même dans tous les énoncés, ce sens étant donc entièrement défini par l'apport propre de l'expression (exemples : *tournevis*, *hectolitre*...). A l'autre extrême, on trouve les homonymes « purs », dont l'apport constant, commun à tous les emplois, est effectivement nul, puisque le sens peut changer radicalement suivant les énoncés (exemples : *avocat*, *vol*...). Entre ces deux extrêmes, se trouve le cas général de la polysémie, avec des cas qui tendent vers la monosémie, quand le contexte ne joue qu'un rôle minime (tous les sens recensables sont très proches les uns des autres), et d'autres vers l'homonymie, quand l'apport propre constant est très faible⁴.

Prenons quelques exemples pour illustrer ce dernier point. Soit le mot *bureau*. Il possède quatre sens principaux : un meuble (ex. : *s'asseoir à son bureau*), une pièce (ex. : *ouvrir la fenêtre de son bureau*), un établissement (ex. : *le bureau de poste*, *le bureau de tabac*, etc.), une institution (ex. : *le bureau de l'Assemblée*, *le bureau de l'association*, etc.). Ces différents sens sont indéniablement reliés, ce qui signifie que l'on a bien affaire à de la polysémie et non à de l'homonymie. Cependant, quand on essaie de déterminer l'apport propre du mot *bureau* qui est commun à tous ses emplois, on s'aperçoit qu'il est très ténu : une vague notion d'activité d'écriture, qu'il semble difficile de rendre opératoire dans un calcul effectif du sens de *bureau* en contexte.

Prenons maintenant le mot *livre*. Ce nom est particulièrement intéressant car il relève d'une polysémie tout à fait particulière, que Kleiber (1999) appelle la polysémie logique, et qui se situe à la frontière entre polysémie lexicale et variation contextuelle. Le problème classique posé par *livre* est qu'il semble présenter deux sens principaux, attestés par les dictionnaires, mais que ces deux sens possèdent des propriétés sémantiques qui empêchent de les considérer comme les différents sens d'un polysème standard. Le TLFi distingue ainsi deux sens pour *livre*. L'un désigne le livre en tant qu'objet, assemblage de feuilles destiné à être lus. Dans ce sens *livre* désigne un objet matériel. L'autre désigne le livre en tant que contenu, l'œuvre en elle-même. Dans ce sens, *livre* désigne un objet abstrait, informatif. On peut voir en *livre* un polysème standard, prenant selon le contexte un sens ou l'autre. Mais la particularité de *livre* (et d'autres noms comme *poulet*, *film*, *banque* ou les noms de ville) est, d'une part, que ses différents sens peuvent se retrouver unifiés dans un sens global, comme dans l'énoncé 3, et que, d'autre part, ils ne sont pas concurrents et peuvent coopérer au sein d'un même énoncé (4 et 5) – ce qui est quasi impossible pour les sens d'un polysème standard.

(3) *J'ai acheté un nouveau livre.*

(4) *Le petit livre tout jauni que tu m'as prêté est particulièrement intéressant.*

(5) *C'est un livre lourd à trimbaler mais très facile à lire.*

On a donc affaire dans ce cas à une polysémie beaucoup plus proche de la monosémie que de l'homonymie. Certains auteurs (Croft & Cruse 2004, Jayez 2008) parlent alors de facettes.

Un des intérêts des espaces sémantiques que nous construisons est de pouvoir rendre compte de ce continuum entre polysémie et homonymie. Cela permet de pourvoir traiter les cas intermédiaires, comme *livre* et *bureau*, sans avoir à décider à l'avance d'un éventuel typage particulier. Nous ne sommes bien sûr pas en mesure de représenter correctement les monosèmes, car ils correspondent à des sens très spécifiques et ne possèdent donc que très peu de synonymes. Notre algorithme ne permet pas de leur associer

⁴ Notre approche de la polysémie est détaillée dans Jacquet et al, 2005.

un espace sémantique. Certains (comme *tournevis*) sont même absents du DES. Nous obtenons en revanche des résultats intéressants quand il s'agit de caractériser différents types de polysèmes. Nous présentons à titre d'exemple les espaces sémantiques associés aux noms *vol*, *bureau* et *livre*. Les espaces sémantiques rendent clairement compte des distinctions que nous venons de mentionner :

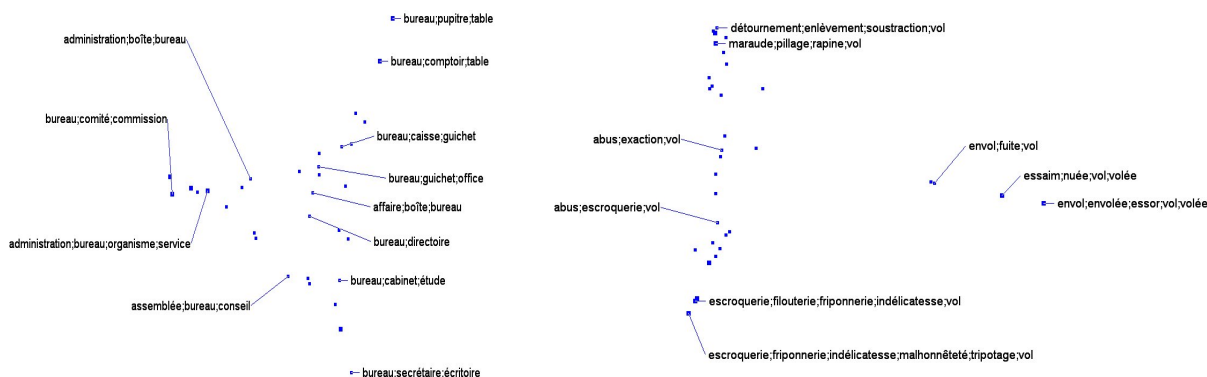


Figure 2. Espaces sémantiques associés à *bureau* et *vol*

Espace sémantique : *livre* (28 unités, 23 cliques) - axes 1 et 2

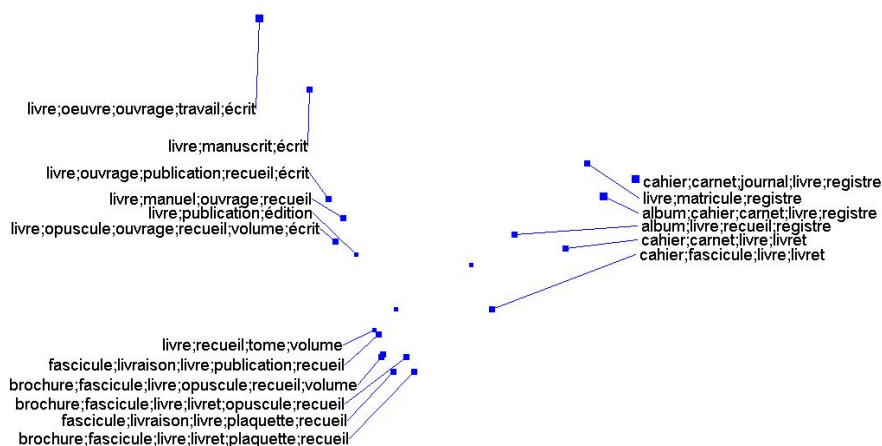


Figure 3. Espace sémantique associé à *livre*

- Figure 2 : L'espace sémantique de *bureau* est clairement continu. Même s'il est difficile à déterminer, le noyau de sens de *bureau* existe et établit des liens sémantiques indéniables entre les différents sens de *bureau*. L'espace sémantique de *vol* est, quant à lui, séparé en deux sous-espaces disjoints, sans chevauchement. On repère ainsi deux lexies bien différentes. On est clairement dans un cas d'homonymie.
- Figure 3 : L'espace sémantique de *livre*⁵ fait apparaître deux branches. La partie supérieure gauche de l'espace sémantique rassemble les cliques contenant les synonymes comme *oeuvre*, *travail*, *écrit*, *ouvrage*, c'est-à-dire les sens correspondant au contenu d'un livre. La partie droite de l'espace sémantique rassemble les cliques contenant les synonymes comme *album*, *cahier*, *registre*, c'est-à-dire les sens correspondant à des emplois de *livre* comme objet matériel, assemblage de feuille et support d'écriture. Ces deux branches se rejoignent au niveau de la partie centrale inférieure de l'espace sémantique, qui constitue une région de sens intermédiaire entre les

⁵ Pour une étude approfondie, voir Venant (2008).

deux précédentes. Les cliques de cette région de l'espace contiennent des procès de type /publication/, comme *livraison*, *plaque*, *brochure*.... L'accent est mis ici sur l'objet que l'on obtient, le résultat de la publication. La topologie obtenue pour *livre* fait donc état d'un continuum de sens, qui permet de passer progressivement des sens /contenu à l'état pur/, à l'extrémité de la branche gauche, aux sens /objet à l'état pur/, à l'extrémité de la branche droite. La continuité est assurée par les sens /résultat d'une publication/, ce qui paraît tout à fait cohérent puisque l'un des objets de la publication, c'est précisément de faire le lien entre le contenu et la forme. Il est difficile de séparer le contenu et la publication, tout simplement parce que le contenu d'un livre est conçu par l'écrivain comme devant être publié, et il n'est accessible au lecteur que parce qu'il a été publié. De même il est difficile de séparer la publication de l'objet matériel puisqu'il en est le résultat.

On voit sur ces quelques exemples que l'algorithme utilisé permet d'obtenir, de façon totalement automatique, et très rapidement, des représentations sémantiques fiables, fidèles à la nature sémantique de l'unité étudiée. Elles rendent compte des distinctions de sens que l'on peut trouver dans les dictionnaires, tout en proposant un éclairage sémantique différent. Ainsi, dans le cas de *livre*, on voit apparaître une zone de sens centrale, le livre comme résultat d'une publication, dont la spécificité est d'être un sens à la fois intermédiaire et synthétique entre les facettes classiques *objet* et *contenu*. Elle exprime une propriété fondamentale du livre, celle qui distingue le livre d'une simple lettre ou d'un journal intime, à savoir la **dimension sociale** du livre. Car le livre est plus qu'un contenu s'appuyant sur une certaine forme. Le livre est avant tout un objet social, doté d'un contenu destiné à être rendu public. C'est en intégrant cette dimension, apparemment oubliée des analyses classiques, que l'on pourra interpréter des expressions comme « un livre à succès », « un livre qui tranche, qui innove », « un livre qui est passé inaperçu. »

6 Zones de sens dans un espace sémantique

Les espaces sémantiques tels que nous les construisons permettent une étude approfondie de la sémantique d'une unité donnée, en mettant au jour une structure sémantique sous-jacente. Ils sont tels quels très utilisés par les internautes s'interrogeant sur les différentes acceptions d'une unité et/ou les synonymes pertinents dans un contexte donné. Ils sont aussi utilisés par les linguistes s'intéressant aux caractéristiques sémantiques des unités lexicales. François et Senechal (2004) les utilisent par exemple pour caractériser les différents foyers de polysémie d'un verbe. Nous avons pour notre part cherché à les exploiter pour modéliser le sens pris par un vocable polysémique dans un énoncé (Venant 2004, 2006, 2008, Jacquet, 2003). L'idée est de modéliser l'interaction entre le noyau de sens et le cotexte, afin de déterminer la zone de l'espace sémantique qui correspond au sens pris par l'unité considérée dans le cotexte étudié. Nous ne détaillons pas ici le calcul effectué. Disons simplement que nous calculons un taux d'affinité entre le cotexte et chaque clique de l'espace sémantique. Ce calcul est effectué à partir de données de cooccurrences issues d'un corpus. Les taux d'affinité calculés permettent d'obtenir une « déformation » de l'espace sémantique. La zone de déformation correspond au sens pris par l'unité étudiée dans le contexte considéré. Le mode de représentation, continu, que nous avons choisi prend ici tout son intérêt car il nous permet de modéliser le sens par une région de l'espace sémantique. Le fait d'utiliser une région, et non un point, permet de rendre compte de tous les cas de figure interprétatifs. Une région étroite correspond à un sens précis, une région étendue à un sens plus indéterminé, une région non connexe à une ambiguïté.

Ce mode de représentation permet ainsi de rendre compte de la polysémie logique⁶. La figure 4 montre des zones de sens calculées sur l'espace sémantique de *livre*. La partie gauche de la figure 4 présente les zones de sens activées par l'adjectif *ancien* (correspondant donc aux sens possibles pour *livre* dans *un livre ancien* ou *un ancien livre*). On y voit deux régions nettement séparées. On est donc en présence d'une ambiguïté. *Ancien* sélectionne tantôt le sens /contenu/ de *livre* (comme dans « J'aime son dernier roman mais je trouve ses anciens livres mieux écrits »), tantôt le sens /objet/ de *livre* (comme dans « J'ai trouvé un joli livre ancien chez l'antiquaire »). C'est le contexte élargi qui permet de trancher.

⁶ voir Venant (2008).

La partie droite de la figure 4 montre les zones de sens obtenues pour l'adjectif *prochain*. On y voit comment l'unification des facettes /contenu/ et /objet/ se fait, via l'activation de la zone centrale de l'espace correspondant au livre en tant que publication. Dans l'expression *prochain livre*, on réfère à la fois à l'objet manufacturé, résultat de la publication, et au contenu, comme en témoigne l'énoncé : « Entre deux livres, et plus précisément dans ce temps où il n'y a encore pas de place bien définie pour cet objet inexistant qu'est « le prochain livre », c'est là que viennent s'accumuler les notations, les intentions[...] » (Genèse du romain contemporain, D.Ferrer et B. Boie, p44)

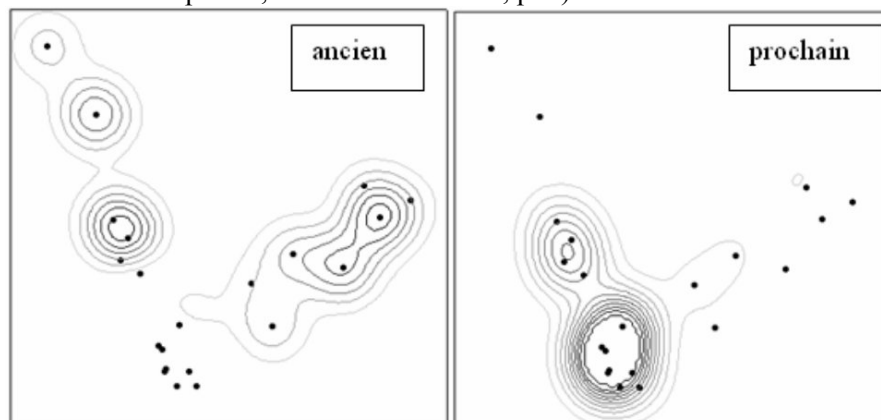


Figure 4. Zone de sens dans l'espace sémantique de *livre*

7 Granularité et changement d'échelle

Notre modèle repose donc sur le choix d'une représentation continue du sens lexical, et en particulier sur le choix des cliques comme unité minimale de représentation sémantique et sur le tuilage continu de l'espace sémantique qu'elles réalisent. Cependant, les cliques représentent des nuances de sens très précises, souvent très proches les unes des autres. Cette subtilité sémantique génère une redondance qui peut poser problème dans certaines applications. Une granularité aussi fine rend difficile, voire impossible, l'étude des relations entre unités lexicales, ainsi que l'articulation entre niveau conceptuel et sémantique. La représentation continue du sens est très adaptée pour certaines tâches, comme l'étude des ambiguïtés ou la recherche d'un synonyme pertinent pour une unité donnée dans un contexte donnée. En revanche, une désambiguïsation à grande échelle, ou un étiquetage sémantique, vont demander l'utilisation d'étiquettes macroscopiques, et le regroupement des cliques en zones de sens correspondant, par exemple, aux définitions d'un dictionnaire. Nous cherchons donc désormais à avancer vers une représentation à granularité variable, où l'on pourrait, selon les besoins, accéder aux nuances de sens représentées par les cliques, ou à des sens macroscopiques correspondant à des regroupements de cliques. Plusieurs pistes sont en cours. Elles visent à repérer des zones denses dans les nuages de points formés par les cliques, soit par des méthodes géométriques de clusterisation, soit par interaction avec d'autres ressources sémantiques.

La possibilité de grouper les cliques permet aussi d'envisager des changements d'échelle. La clique constitue en effet une unité structurale trop fine pour une étude globale des relations sémantiques au sein du lexique. Le changement de granularité permet de prendre du recul et d'étudier la place des unités lexicales dans l'organisation globale du lexique.

7.1 Un espace global

La représentation que nous venons de décrire est centrée sur une seule unité linguistique, par exemple le nom *cours*. Les synonymes de *cours* n'interviennent qu'en tant qu'ils permettent de distinguer ses différentes acceptions : ils restent cantonnés à l'intérieur de l'espace sémantique de *cours*. Les seuls sens de ces synonymes qui sont pris en considération sont ceux qu'ils partagent avec l'unité sur laquelle on s'est focalisée. Il s'agit donc d'une représentation tronquée de la relation de synonymie, puisque seul l'un des synonymes (en l'occurrence *cours*) se trouve pleinement représenté. Pour représenter plus fidèlement les

relations entre *cours* et ses synonymes, il faut donc passer d'un espace sémantique *local*, associé à une unité, à un espace sémantique *global*, qui replace l'ensemble des unités concernées dans le réseau complet ou dans un sous-réseau paradigmatique. D'une manière générale, ce n'est que dans un espace global que l'on peut obtenir une représentation fidèle des relations de synonymie entre plusieurs unités lexicales.

La clique a montré son efficacité en tant qu'outil d'exploration d'un graphe lexical. Le problème, on l'a dit est celui du niveau de granularité. Si on travaille, par exemple, sur un graphe adjectival⁷, on se retrouve avec plus de 11000 cliques. Un si grand nombre de cliques ne permet d'envisager ni l'exploration, ni l'exploitation, d'un tel espace sémantique. L'idée est donc de repérer des branches de cliques correspondant à des sens très proches mais pas forcément deux à deux. Pour les mettre en évidence, nous avons défini un outil géométrique, la **boule**, à partir duquel nous définissons et construisons des **branches** de cliques. On forme autour de chaque clique une boule contenant ses voisines les plus proches, à l'exclusion de celles qui n'ont aucun synonyme commun avec elle. Les branches que nous cherchons à mettre en évidence sont alors des rassemblements de boules, rassemblées en fonction du nombre de synonymes qu'elles ont en commun.

La mise en œuvre de ces outils sur un graphe adjectival (Venant, 2006) nous a permis d'une part de caractériser les grandes classes adjectivales traditionnellement distinguées : qualificatifs, intensifs (Romero 2004), relationnels (Daille 2001), primaires (Borodina 1963, Goes 1999, Noailly 1999), et d'autre part de montrer que, d'un point de vue théorique, il ne fallait pas chercher à classer les adjectifs eux-mêmes, mais leurs emplois, un même adjectif pouvant appartenir à différentes classes suivant ses emplois. Ainsi nous avons pu montrer que la plupart des adjectifs, même les plus qualificatifs, possèdent des emplois intensifs, comme par exemple *méchant* dans des emplois tels que *une méchante voiture* (intensif positif) ou *un méchant costume de laine* (intensif négatif). Il existe d'autre part un continuum entre les différents types d'emplois adjectivaux, qu'on se situe au niveau d'un adjectif donné comme *intime* : de *l'ami intime* (qualificatif) à *l'intime conviction* (intensif), ou encore *adolescent* : de *l'amour adolescent* (qualificatif) à *un groupe adolescent* (relationnel), ou que l'on prenne le lexique dans son ensemble : des adjectifs qualificatifs intenses comme *bouillant* (*eau bouillante*) aux intensifs purs comme *énorme* (*énorme envie*). Ceci explique pourquoi certains linguistes, comme Bartning & Noailly (1993), ont tant peiné en cherchant à établir des frontières nettes, notamment entre adjectifs qualificatifs et adjectifs relationnels. Nos résultats nous amènent plutôt à adhérer à la proposition de Goes (1999) ou de Romero (2004) de parler d'adjectifs « statistiquement relationnels » (ou « statistiquement intensifs ») pour caractériser des adjectifs comme *procédural* (resp. *extrême*), dont les emplois sont majoritairement relationnels (resp. majoritairement intensifs), sans que cela exclue la possibilité de trouver ces adjectifs dans des emplois purement intensifs, ex. : *une lenteur procédurale* (resp. purement qualificatifs, ex. : *une expérience extrême*).

7.2 Vers une représentation conceptuelle

Une autre piste pour représenter les sens macroscopiques d'une unité lexicale est d'apparier les cliques d'une unité donnée aux descriptions sémantiques d'une approche discrète. Nous avons initié le travail en utilisant les définitions du TLFI. Pour cela, nous exploitons le travail réalisé par Falk et al (2009) qui vise à apparier, via un calcul de similarités, les sens possibles d'un lexème avec les ensembles de synonymes appropriés. L'idée est de projeter sur les cliques les appariements synonymes/définitions ainsi obtenus. Falk et al se sont pour l'instant limitées au lexique verbal.

A titre de première expérimentation, nous avons calculé pour chaque clique du verbe *abandonner* un taux d'affinité avec une définition du TLFI. Le calcul est le suivant: soit *c* une clique et *d* une définition, on note *Sc* l'ensemble des synonymes appartenant à la clique *c*, et *Sd* l'ensemble des synonymes appariés à la définition *d* par Falk et al. Le taux d'affinité *T(c, d)* entre la clique *c* et la définition *d* est défini par $T(c,d) = |Sc \cap Sd| / |Sc|$

La Figure 5 montre les projections obtenues automatiquement : dans la partie gauche, on a marqué d'un astérisque les cliques ayant un taux d'affinité supérieur à 60% avec la définition «Renoncer à un pouvoir,

⁷Nous avons extrait du DES un graphe adjectival. Il comporte ses 198 549 arcs pour 49133 sommets et contient 11900 cliques.

à des droits, à la possession d'un bien ou à l'utilisation d'une chose ». Dans la partie droite, on a marqué les cliques ayant un taux d'affinité supérieur à 60% avec le sens « Quitter un lieu. »

On voit que l'utilisation du taux d'affinité peut constituer un bon moyen de repérer automatiquement dans le nuage de points formé par les cliques, les sens macroscopiques correspondant aux définitions du TLFi. On peut ainsi espérer structurer de façon automatique les espaces sémantiques, repérer et caractériser automatiquement les différentes lexies d'un polysème et enrichir notre représentation sémantique avec les informations spécifiées dans les définitions du dictionnaire (domaine, constructions, sous-catégorisation, combinatoire...).

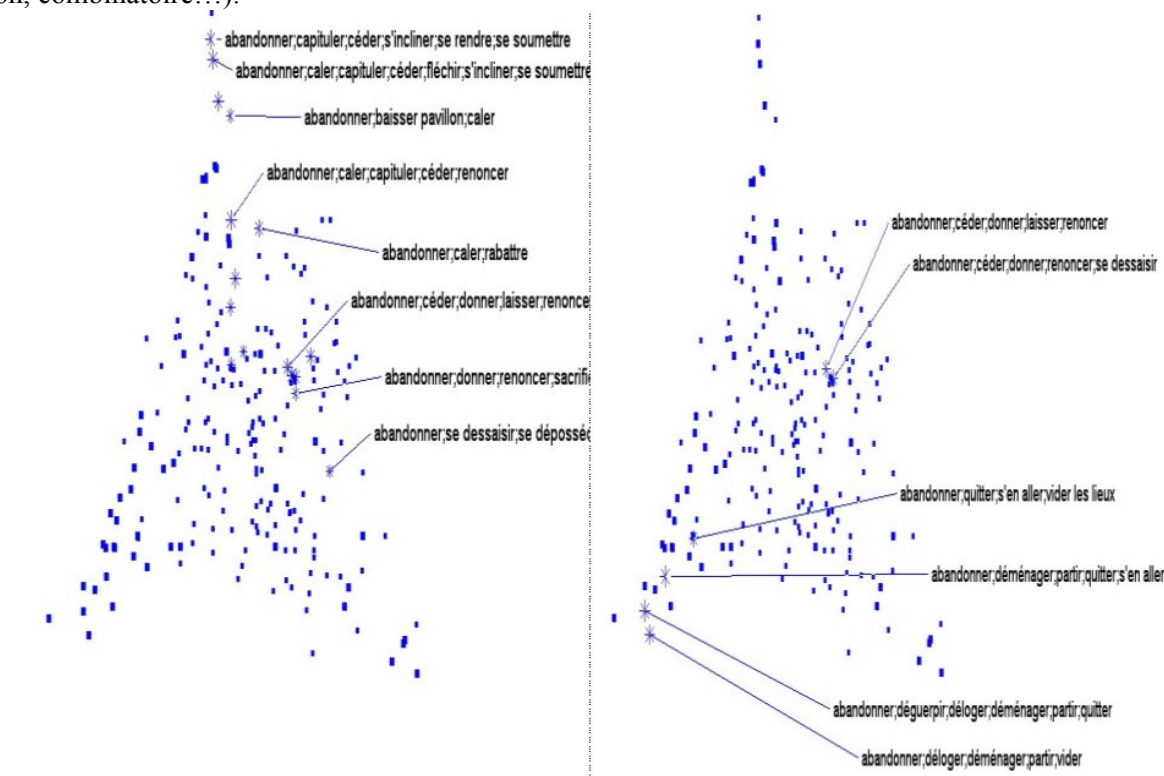


Figure 5. Projection de définitions du TLFi sur l'espace sémantique de *abandonner*

8 Conclusion

Nous avons présenté ici un modèle de représentation continu du sens. Ce modèle présente l'avantage d'être adapté à la fois à des utilisations humaines et automatiques. Le choix de la relation de synonymie comme accès aux informations sémantiques lexicales s'est révélé pertinent pour la caractérisation sémantique des polysèmes. La question qui se pose désormais est celle de l'enrichissement et de la structuration de la ressource sémantique ainsi obtenue. Nous ne rendons compte pour l'instant que de certains aspects de la structure lexicale, puisque nous ne l'avons abordée que par le biais des relations de synonymie. Cette vue, partielle, est à compléter par l'étude d'autres relations lexicales qu'elles soient sémantiques (antonymie, hyperonymie...) ou non (dérivation, suites syntaxiques : adjectif – nom, verbe – adverbe..., rapports syntaxiques : verbe – nom (sujet), verbe – nom (objet)...). Un pas vers dans cette direction a déjà été fait avec la construction d'espaces distributionnels pour étudier la polysémie verbale (Jacquet & Venant, 2005). La prochaine phase consistera à tenter d'explorer d'autres graphes lexicaux tout en travaillant à stabiliser les modes de construction et d'exploration des espaces globaux. Les pistes envisagées, clustering sur les cliques, ou interaction avec d'autres modèles sémantiques semblent prometteuses.

References

- Croft, William & D.Allan, Cruse. 2004. *Cognitive Linguistics*, Cambridge University Press.
- Bartning, Inge. & Michèle, Noailly. 1993 du relationnel au qualificatif : flux et reflux, *L'information Grammaticale*, 58, L'adjectif (M. Noailly ed.).
- Borodina, M.A. 1963. L'adjectif et les rapports entre sémantique et grammaire en français moderne., dans *Le Français Moderne*, XXXI-3, p. 193-198., 1963
- Daille, Béatrice. 2001. « L'identification en corpus d'adjectifs relationnels : une piste pour l'extraction automatique de terminologie », *TAL, Volume 42 Lexiques sémantiques*
- Falk, Ingrid, Claire, Gardent, Jacquy, Evelyne & Fabienne, Venant (à paraître). A method for grouping synonyms, Lexicography in the 21st century: new applications, new challenges. Louvain la neuve.
- François, Jacques & Morgane, Sénéchal. 2004. Le sémantisme propre des cadres prédicatifs et la polysémie des verbes de parole, actes du colloque *La prédication*, Aix-en-Provence.
- Goes, Jan. 1999. *L'adjectif entre nom et verbe*, Paris – Louvain –La - Neuve, Duclot.
- Jacquet, Guillaume. 2003. Polysémie verbale et construction syntaxique : étude sur le verbe jouer, *Actes de la 10ème conférence annuelle sur le Traitement Automatique des Langues (RECITAL 2003)*, Batz-sur-mer.
- Jacquet, Guillaume & Fabienne, Venant. 2005. Construction automatique de classes de sélection distributionnelle, Actes de la 10^{ème} conférence annuelle sur le Traitement Automatique des Langues, TALN 05. Dourdan.
- Jacquet, Guillaume, Fabienne, Venant & Bernard Victorri. 2005. Polysémie lexicale, dans ENJALBERT P. *Sémantique et traitement automatique du langage naturel*, Hermes.
- Jayez, Jacques, 2008. Quel rôle pour les facettes ? *Langage*, volume 172.
- Kleiber, Georges. 1999. *Problèmes de sémantique - la polysémie en questions*, Presses universitaires du Septentrion, Villeneuve d'Ascq
- Noailly, Michèle. 1999. *L'adjectif en français moderne*, Paris, Ophrys.
- Ploux, Sabine & Bernard Victorri. 1998. Construction d'espaces sémantiques à l'aide de dictionnaires de synonymes, *Traitement automatique des langues*, 39/1, p.161-182, 1998
- Polguère Alain. 2002. Le sens linguistique peut-il être visualisé ? In D. Lagorgette et P. Larrivée (dir.) : *Représentations du sens linguistique*, coll. "Lincom Studies in Theoretical Linguistics", 25, Munich : Lincom Europa, 89-103
- Pustejosky, James. 1995. *The generative lexicon*, Cambridge, MIT Press, 1995.
- Romero, Clara. 2004. « Les adjectifs intensifs », *François J. l'adjectif en français et à travers les langues*, Presses Universitaires de Caen, 2004.
- Venant, Fabienne. 2004. Polysémie et calcul dynamique du sens, *Le poids des mots, actes des JADT 04*. Louvain La neuve.
- Venant, Fabienne. 2006. Représentation et calcul dynamique : exploration du lexique adjectival du français, mémoire de doctorat de l'EHESS.
- Venant, Fabienne. 2008. Représentation géométrique et calcul dynamique du sens lexical : application à la polysémie de livre. *Langages*, 172.
- Victorri, Bernard. 1997. La polysémie : un artefact de la linguistique ? *Revue de Sémantique et de Pragmatique*, 2, p. 41-62.
- Victorri, Bernard & Catherine, Fuchs. 1996. *La polysémie, construction dynamique du sens*, Paris, Hermès.